Heating Pattern Cluster Analysis of Winter AMI Data

Efficiency Vermont R&D Project: Greenhouse Gas Reduction

October 2024

Xavier Hayden
Abigail Hotaling

Efficiency Vermont

20 Winooski Falls Way
Winooski, VT 05404

Contents

Executive Summary

As electrification remains a key strategy for reducing greenhouse gas emissions, Efficiency Vermont is focused on furthering electrification through energy efficient fuel switches and upgrades. One challenge identified through partnership with distribution utility partners is identifying customers relying on unregulated fuel for heating and cooling of homes. Intentional approaches are necessary to transition consumers from unregulated fuel markets to efficient electric heating and cooling systems, especially in the face of rising fuel costs and maintenance expenses. Identifying these consumers poses a challenge due to the lack of reporting standards for delivered fuel suppliers compared to electric utilities. Efficiency Vermont can leverage AMI data to identify customers using non-electric heating by analyzing energy use patterns relative to outdoor air temperature. For example, customers that use non-electric sources for heating would show little to no correlation of electric consumption during low outdoor air temperatures. Customers that display this relationship with electric consumption and outdoor temperature are ideal participants for electrification programs.

To identify target markets, methods analyzing Advanced Metering Infrastructure (AMI) and outdoor temperature data, offer a promising solution. Using open-source regression models, Efficiency Vermont can disaggregate heating and cooling loads for individual customers. This allows for Efficiency Vermont to create a detailed landscape of electric heating and cooling usage for a given population, allowing for quantitative comparisons. This paper applies k-means clustering to residential AMI data to distinguish sub-populations based on their heating sources, focusing on identifying those reliant on non-electric heating. By determining this group, Efficiency Vermont can design more intentional electrification programs to achieve carbon reduction goals.

This project set out to create a data-driven process to unveil unique energy use patterns in a population. This paper used 7,430 Vermont Electric Coop residential sites as a total sample size. While each category can have unique properties that warrant their own use-cases, one group should have features best explained by the absence of electric heating. The results show six distinct clusters, each representing unique energy usage profiles. Of these groups, three showed similar profiles with varying degrees of heating and cooling usage, in total making up around 66% of the population sample. Another group which represented 8% of the sample population showed virtually no cooling usage at all. Lastly, two groups stood out as the most likely to use non-electric heating, making up 26% of the total sample population and showed minimal electrical heating at cool temperatures. This is indicated by the extremely low heating balance point which implies that residences in these groups are not using electricity as their main source of heating. These findings highlight significant opportunities to optimize and support energy efficiency programs, particularly for homes where electrification could make the most impact.

By incorporating more focused targeting through creating subgroups based on energy disaggregation, Efficiency Vermont can gain valuable knowledge of residential buildings that use non-electric heating. This knowledge can help transition customers away from unregulated fuel markets and towards efficient heating and cooling systems. Fossil fuels, for example, can have higher greenhouse gas impacts than other unregulated fuels. By transitioning to their electrified counter parts, this could lead to higher greenhouse gas reductions as well as cost savings for the customer.

Limitations in this work, such as identifying the unregulated fuel being used among residential sites, keep the scope of this study focused on the regulation fuel status, rather than identifying specific fuel type. Additionally, the lack of verifiable data to measure model accuracy effect the current, applicable scope of this analysis. Future work could combine AMI data with more granular data sources to further enhance the identification of non-electric heating users. For example, performing a deeper analysis on customers with known heat pump installations, could help distinguish heat pump performance and optimization for a given population.

Introduction

Energy efficiency continues to be a leading approach in reducing greenhouse gas emissions. EVT has already explored s electrification projects based solely on the use of AMI data. Therefore, exploring new methodologies while bringing in additional datasets allows us to ask new questions, such as how can we identify opportunities to transition customers from unregulated fuels markets to efficient electric heating and cooling technologies? Consumers of unregulated fuel are particularly difficult to identify, as delivered fuel suppliers do not have to meet the same reporting standards as electric distribution utilities. With the lack of complete data about the market, energy efficiency programs need alternative methods to identify this customer base.

Efficiency Vermont can leverage Advanced Metering Infrastructure (AMI) data to distinguish groups of customers in a population, in order to focus program design and customer outreach. AMI consists of a network of meter communication and data management that collects, transmits, and stores electrical consumption data—and has established itself as a cornerstone in the modernization of utility systems. The high-frequency interval data recorded from AMI meters enables more granular disaggregation of energy use into heating, cooling, and time-of-use patterns.

The objective of this research is to leverage residential AMI data and outdoor air temperature to determine, among a base of residential customers, which residences are heating their homes primarily through non-electric means. Determining this subpopulation will allow Efficiency Vermont the ability to select groups of customers in higher need of electrification upgrades more effectively.

Electrical demands generally increase as temperatures rise above one threshold or drop below another threshold. A popular way to model this type of behavior is to use a U-shaped segmented regression model, which can identify transitions between heating and cooling loads as seen in Figure 7. Shifting the focus away from time-of-day usage and towards daily change in average net power usage enables drawing further insights into what indicators challenge a customer building’s energy retention.

Background

Ideally, energy program outreach starts with identifying customers who have the highest potential for energy savings based on current operating practices and equipment. Efficiency Vermont’s previous work has demonstrated the use of AMI data to identify customers with high savings or demand-management potential based on energy usage characteristics and load shape patterns in both the residential and commercial sectors. Investments in energy efficiency are often most effective for high-use industrial and commercial customers; programs for residential customers face a natural challenge in scale. In some scenarios, it is not efficient to analyze energy usage and savings for each individual residence; instead, we can use automated machine-learning analysis to identify savings potential for groups of customers who individually might struggle to achieve priority recognition for energy efficiency programming. We can implement this process to recommend a variety of programs across populations of interest or track these analytics wholesale in our ingest pipelines.

Building on Efficiency Vermont’s previous work, this research proposes an analysis technique for residential AMI data, using the relationship between energy usage and outdoor air temperature to disaggregate heating and cooling consumption. Once refined, we can automate this process for use with new residential sites that enter our customer base.

AMI data provide valuable insight into a customer’s energy usage, but do not specify building interior properties or equipment. Combining customer AMI data with outdoor air temperature can help describe the relationship between a building's energy usage and changing ambient temperatures. Disaggregating energy data into baseline load, heating load, and cooling load can indicate if there is electric heating, ventilation, and air conditioning (HVAC) equipment in the building and how it is being used. These data present a unique opportunity to surface higher-level insights into population energy-use trends. Identifying energy savings or transformation opportunities based on identified AMI usage characteristics of customer subpopulations can provide more focused outreach for programs. This allows for energy efficiency implementers to identify customers that may yield the most impact from electrification or energy efficiency measures.

While other work has been done on the disaggregation of heating and cooling energy to determine buildings that primarily use unregulated fuel sources, this often involves using available data on building envelopes. Typically, this information is not accessible on a large scale and can be vulnerable to unreported on-site changes in the building’s envelope. AMI data provide the most insight when paired with complementary customer-related data sources. Widely accessible data that are similarly granular to shorter-frequency AMI data would be ideal for grouping customer subpopulations according to energy-use patterns.

Methods

The focus of this research is buildings' use of energy and which temperatures cause excess energy demands. Defining buildings according to energy model characteristics, or model shape, helps us group them into categories. We can then average each category's constituent models to create a high-level representation of the group at large. Each residence now has a categorical label that describes its heating and cooling energy signatures.

Data Acquisition

This study examines data for residential customers within Vermont Electric Coop (VEC) territory between the years 2021 and 2023. VEC serves around 120,000 customers, with AMI data recorded every quarter-hour to every hour. There were 17,787 active VEC residential utility meters within our date range of interest. The research team began with a random sample of 10,000 residences; the process of building energy models for this population led to an eventual final sample of 7,430 residences with sufficient data (as defined in the following section) to justify the use of our modeling techniques.

We paired each residence with outdoor dry-bulb temperature sourced from Open Weather Map[1]. While geographically specific weather for each VEC residence in our sample would be ideal, due to time and budget constraints, we sourced all weather data for this analysis from Newport, Vermont. Despite natural climate shifts and enhanced greenhouse gas effects, Northern Vermont's climate is still homogenous enough within VEC's territory, which spans a region of no further than a 60 radius, when centered on Newport. Since inter-site climate differences are also not independent of one another, Open Weather Map reports weather at the hourly level, the recommended temperature frequency to use with the modeling software. We selected the years 2021 and 2022 for data availability and completeness.

We based our energy model on normalized total net energy consumption per day (kWh), where net energy consumption is the energy a site consumes from the grid minus the energy it generates. Instead of using the actual net energy consumption per day, our model takes in the net energy consumed that day divided by the total annual usage of net energy consumption at that site. This process scales the data of each site to a common range, allowing for a comparison based on temperature-energy correlation instead of magnitude of energy consumption between different sites. This normalized data allows us to compare relative electric use between sites by measuring the percentage of annual consumption each site uses on a given day. The research team chose to focus on the reported net energy consumption of the building, since our interest lies in the amount of energy that a home is using, rather than the rate at which it uses that energy.

[1] Contains information from VEIC's WEATHER database, which is made available here under the Open Database License (ODbL). Weather data provided by Open Weather.

CalTRACK Daily Model

CalTRACK is a set of methods from the open-source toolkit OpenEEMeter used for estimating avoided energy use (AEU) that relate to the implementation of efficiency measures used on energy-efficient retrofits. These methods yield whole-building, site-level savings outputs and are primarily used in energy efficiency procurement. One of the specifications for this modeling library is the requirement that no less than 10% of the data can be missing, null, or zero. This eliminates part-time residences, like a summer home, so that the model contains only buildings that are in use year-round. The model characterizes each building as a combination of three parts:

- a base load, which refers to the minimum level of energy consumption a building uses independent of weather conditions and seasonal variation (examples of end uses included in the base load are refrigeration and lighting);
- a heating load, which represents the energy required to maintain the interior climate as outdoor air temperatures fall; and
- a cooling load, which represents the energy required to maintain the interior climate as outdoor air temperatures rise.

A building is modeled to determine their base load, heating load, and cooling load. We assume a linear relationship between heating/cooling load and heating/cooling demand, as approximated by heating and cooling degree days, outside of the heating and cooling balance points. The heating balance point is the point at which the model starts to increase its energy usage due to using heating. Likewise, the cooling balance point is the temperature which causes the residence to start to dedicate energy towards cooling.

The model requires daily or hourly temperature data and daily energy usage, building a representative model that describes the temperatures at which a building begins to increase energy for heating or cooling as outdoor temperature increases or decreases. We measure these increases or decreases from heating/cooling balance points. The result is a five-parameter linear equation (1) made up of a baseline intercept (μ_i), the heating and cooling balance points (HDD_p, CDD_p) and the heating and cooling coefficients ($\beta_{H,i}, \beta_{C,i}$)

$$UPD_{p,i} = \mu_i + \beta_{H,i} \cdot HDD_p + \beta_{C,i} \cdot CDD_p + \varepsilon_{p,i} \tag{1}$$

Where ($\varepsilon_{p,i}$) represents an error term to prevent overfitting.

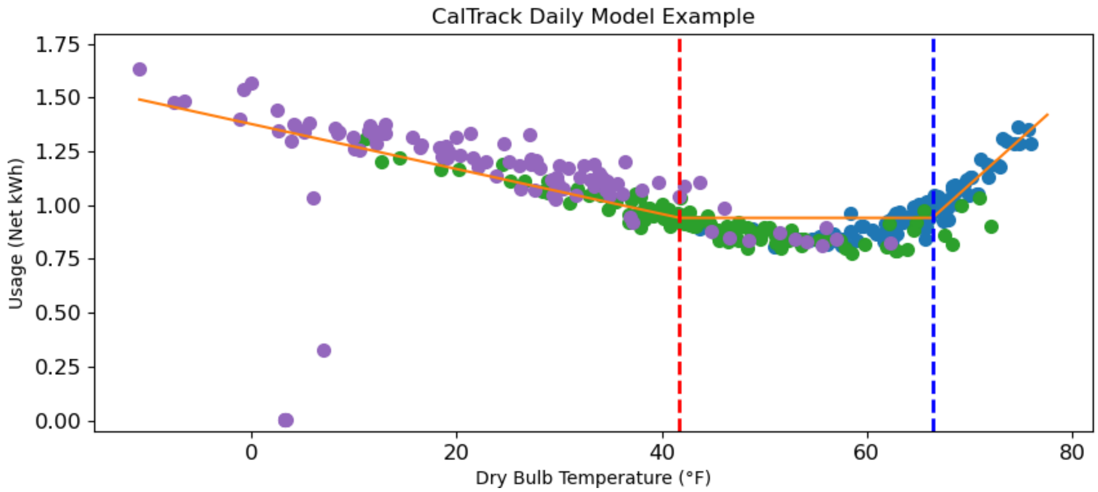

Figure 1. An example of a CalTRACK Daily Model. The points represent the 365 days of a year, while the colors represent winter (purple), shoulder (green), and summer (blue). Shoulder includes both spring and autumn.

The "U" shape that the CalTRACK model takes on allows for the shape to act as a holistic view of the residence's daily heating signature Creating these models for every residence in our sample quantifies the shape of each building's energy signature with five parameters. This quantitative measure allows for classification algorithms, such as K-means clustering, to divide our sample into several distinct groups that have their own unique properties.

Daily Model Clustering

Clustering is a machine-learning technique that groups data into subsets, or clusters, based on their similarities. The goal of clustering is to ensure that data points within a cluster are more similar to each other than to the points in other clusters—like classifying the songs in a music playlist by genre. This technique excels at identifying patterns and provides insights into overarching differences between data.

One of the most common and popular methods of clustering is K-means clustering. This method divides a data set into a predefined number of clusters, *k,* by iteratively assigning each data point to one of the *k* clusters based on the proximity of that data point to the cluster's centroid. The centroid is the average of all the data points assigned to a specific cluster, which the algorithm recalculates in each iteration by reassigning data points to the closest centroid. This process of choosing a centroid, computing and comparing the distance of data of that centroid to the other centroids, is repeated until the distance starts to converge (gains from reassigning data points to centroids give diminishing returns).

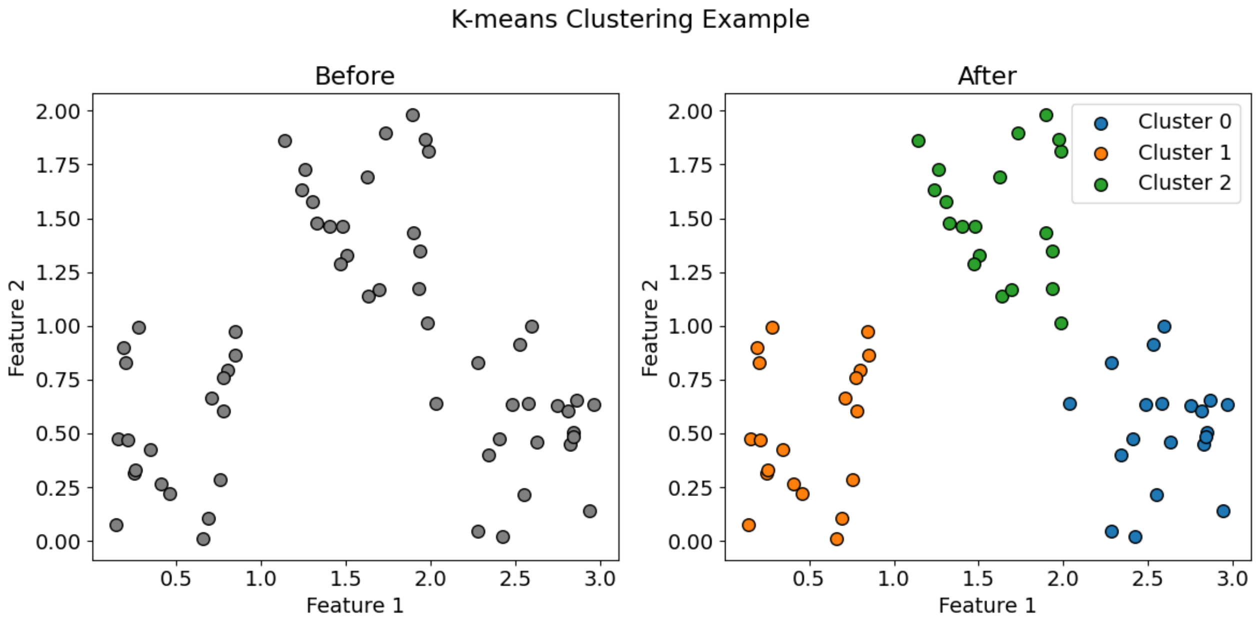

Figure 2. Example K-means clustering algorithm on a two-feature data set.

In the case of using K-means on the residential model parameters accumulated from the previous section, clustering will act on our entire sample of models and classify them based on their parameter values. This produces a list of residences, each classified into a specific group. More importantly, each group contains a centroid that represents the center (average) of the points within. These centroids are the focus of the K-means output and represent the average model parameters within a group of models.

Ideally, we would be able to take the averages for each parameter within each group (i.e. the average heating slope of residences belonging to a specific cluster) and build a new CalTRACK model with those average parameters, but CalTRACK does not currently provide this functionality. To work around this limitation, we averaged the input data of each individual model in a group, producing a final data set that represents the average daily energy usage of models belonging to that group. Building a new daily model with this averaged data set recreates the group's centroid as a CalTRACK model itself, which can give insight into the model features that distinguishes this cluster from the others.

K-means relies on a pre-determined number of clusters to be specified, but there is some nuance in the question of the "best" number of clusters. To choose the number of clusters that would provide the best representation for this project, we use a variety of machine learning algorithms, as well as consideration for industry interests in the results of this work. We evaluate our cluster numbers using *inertia*, a measurement of how well data points fit within their own cluster compared to others. A good classification model strikes a balance between low inertia and a low number of clusters, according to the modeler's needs.

Our first optimization procedure is an elbow analysis. Elbow analysis involves recreating and adding on to the number of clusters in our K-means algorithm and recording the inertia of the new number, repeating the process until the maximum number of clusters has been modeled. Elbow analysis recommends choosing the number of clusters at or around the natural 'elbow' in the graph.

For models ranging from two to fifteen clusters, we also use silhouette analysis. Silhouette analysis tests the distance between points of a cluster and its neighboring cluster. The silhouette coefficient is a way to indicate how far clusters lie from each other. In other words, it measures cluster density and separation. The goal is to choose clusters that have a higher silhouette coefficient, where each cluster is above the average score of all points. Again, this optimization tool requires a degree of nuance depending on the project needs. It is important to select the number of clusters based on models whose silhouette coefficients show evenly measured clusters, with each above the average score—but this is in no way an empirical choice, as it might not capture the modeler's particular needs.

Alongside the silhouette analysis, we plot each cluster in two dimensions. Because the data is being classified based on five distinct features, our data have five dimensions and plotting data in more than three dimensions is a hefty challenge. We use Principal component analysis (PCA), which relies on singular value decomposition, to project higher-dimensional data to a lower-dimensional space. In other words, with minor information loss occurring, we can reduce our five modeled features down to just two engineered features. Plotting in two dimensions gives us visuals to interpret how these clusters relate spatially to one another.

Results

Optimizing Number of Clusters

To narrow down the range of clusters to choose from, we first perform an elbow analysis on models ranging from two clusters up to fifteen. Figure 3 shows the result of this analysis. Where the graph starts to bend and form a natural "elbow," there is a balance between a high enough number of clusters to be descriptive and a reduction in inertia with each additional cluster. In this case, the natural bend in the graph seems to lie between four and six clusters, suggesting that our ideal choice lies in this subrange.

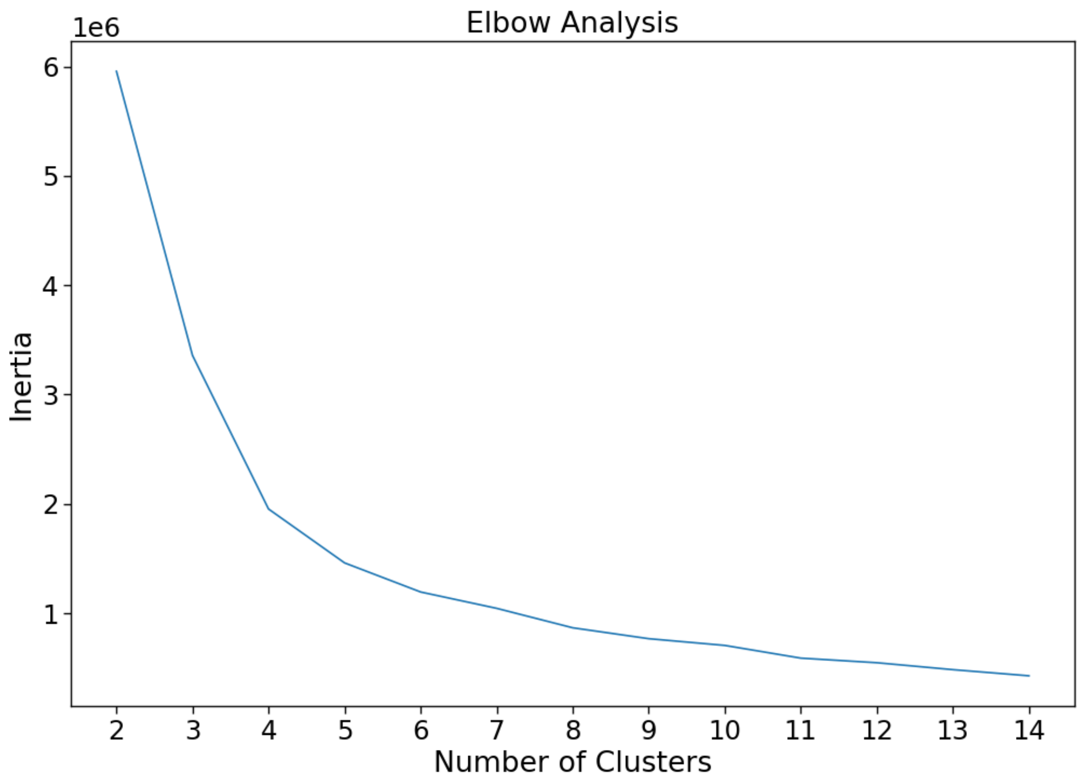

Figure 3. Elbow analysis of model inertia. Note the natural "bend" in the graph that occurs between four and six clusters.

Now that the range of optimal number of clusters has been reduced, we can further test the remaining classifiers by computing a silhouette analysis for model classification of four, five, and six clusters, respectively. For each of these analyses, the signs of a good classification would show clusters of even sizes, where each cluster lies above the average silhouette coefficient value of the whole population. For each of our analysis, all choices of number of clusters produced silhouette coefficient values above the population average, indicating that points were well-defined within their cluster.

The largest cluster in the four-cluster analysis, Cluster 0, holds a much larger number of points compared to the remaining clusters. This indicates a large bias for this cluster that would benefit from being sub-sectioned into more clusters. To capture less obvious groupings, the solution is to add more clusters.

We applied PCA to the data set and plotted alongside the silhouette analysis to give a sense of perspective for how the choice of clusters affects the classification of points. Notice how Group 0 gets split into three when comparing the analysis to classifications of five or six clusters.

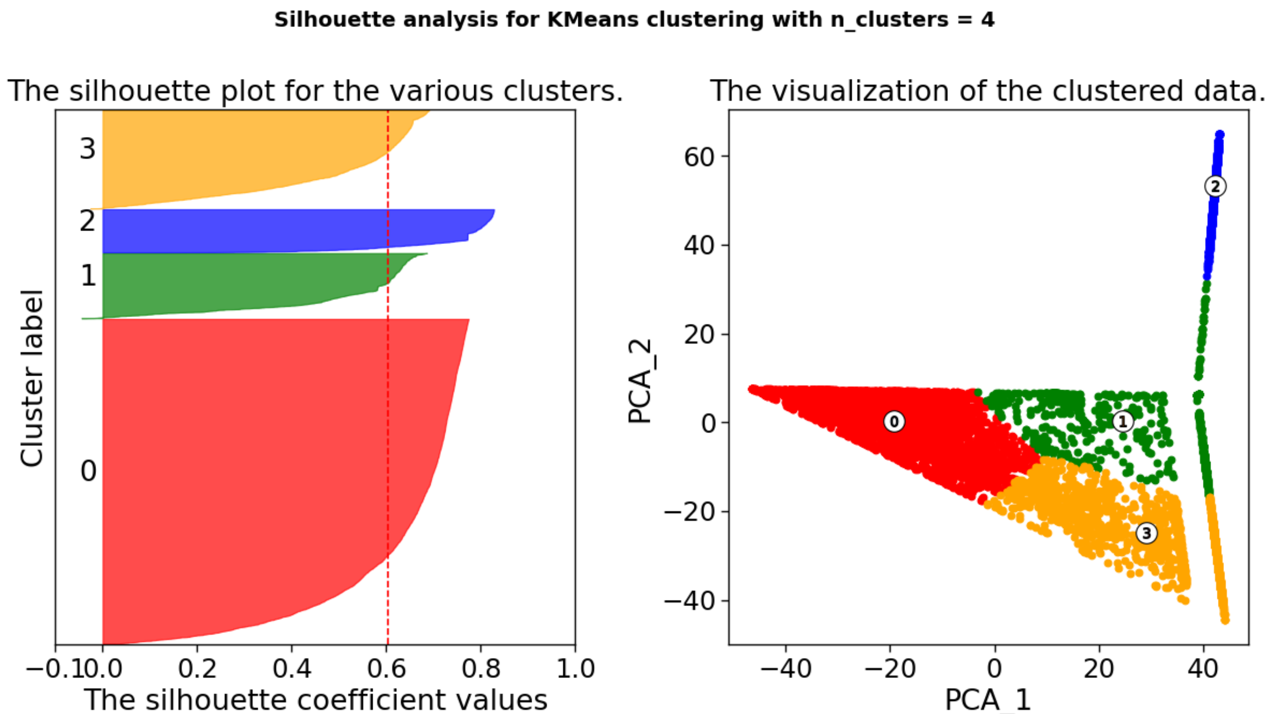

Figure 4. Silhouette analysis for K-means classification with four clusters. The wide fluctuation for Group 0 in the left graph indicates this cluster is overpopulated and would benefit from being broken up into more clusters.

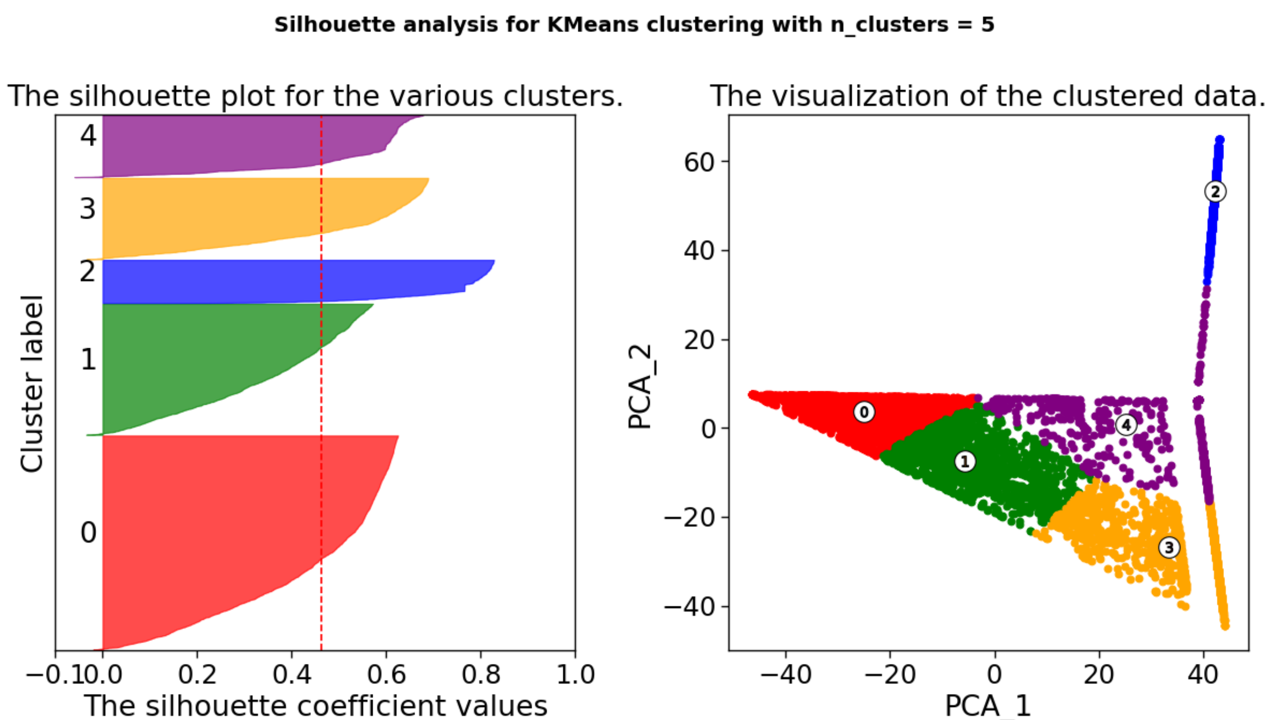

Figure 5. Silhouette analysis for K-means classification with five clusters. The clusters become closer but still have a larger-than-ideal population range.

The silhouette analysis reveals a lot more information about the makeup of our clusters, while the PCA visualization of the cluster space helps visualize the changes K-means creates when adding or subtracting from the number of clusters. While some information loss occurs with dimensionality reduction during PCA, the visualization of these groups provides much more interpretable context.

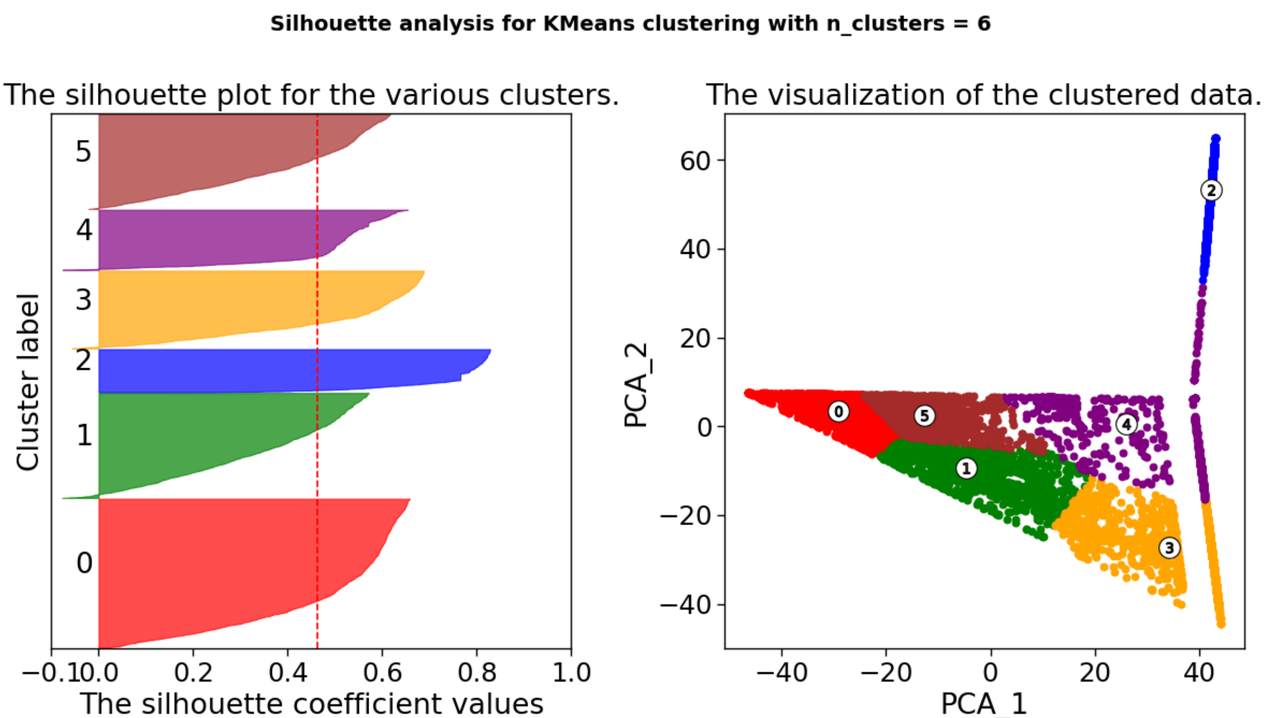

Figure 6. Silhouette analysis for K-means classification with six clusters. Groups are much more evenly populated, and each cluster is well above the average silhouette score.

The silhouette analysis for a six-cluster K-means classifier shows an even distribution of points, with each group having an above-average silhouette coefficient. This points to a strong option in our choice of model. While increasing the number of clusters could improve the analysis, the minor gains through more clustering would be offset by the cost of model-building and risk losing interpretability.

Clustering Results

The results of classifying 7,430 residential CalTRACK models into six groups using K-means clustering can be seen in Figure 7, with the corresponding model values in Table 1. The underlying grey lines on each plot represent the individual CalTRACK models that make up a single group, or cluster. The thicker orange line represents the CalTRACK model created from the average values of the data within each cluster. The dotted red line represents the heating balance point, while the blue represents the cooling balance point.

The five values in Table 1 describe the nature of each cluster. The baseline intercept indicates the residence energy usage when no energy is being dedicated to heating or cooling. Capping either end of this line are the heating and cooling balance points, enclosing the range of temperature values at which this base load occurs. Finally, the rates at which energy usage changes with temperature, which homeowners could view as differing rates of energy efficiency, are described by the heating and cooling slope.

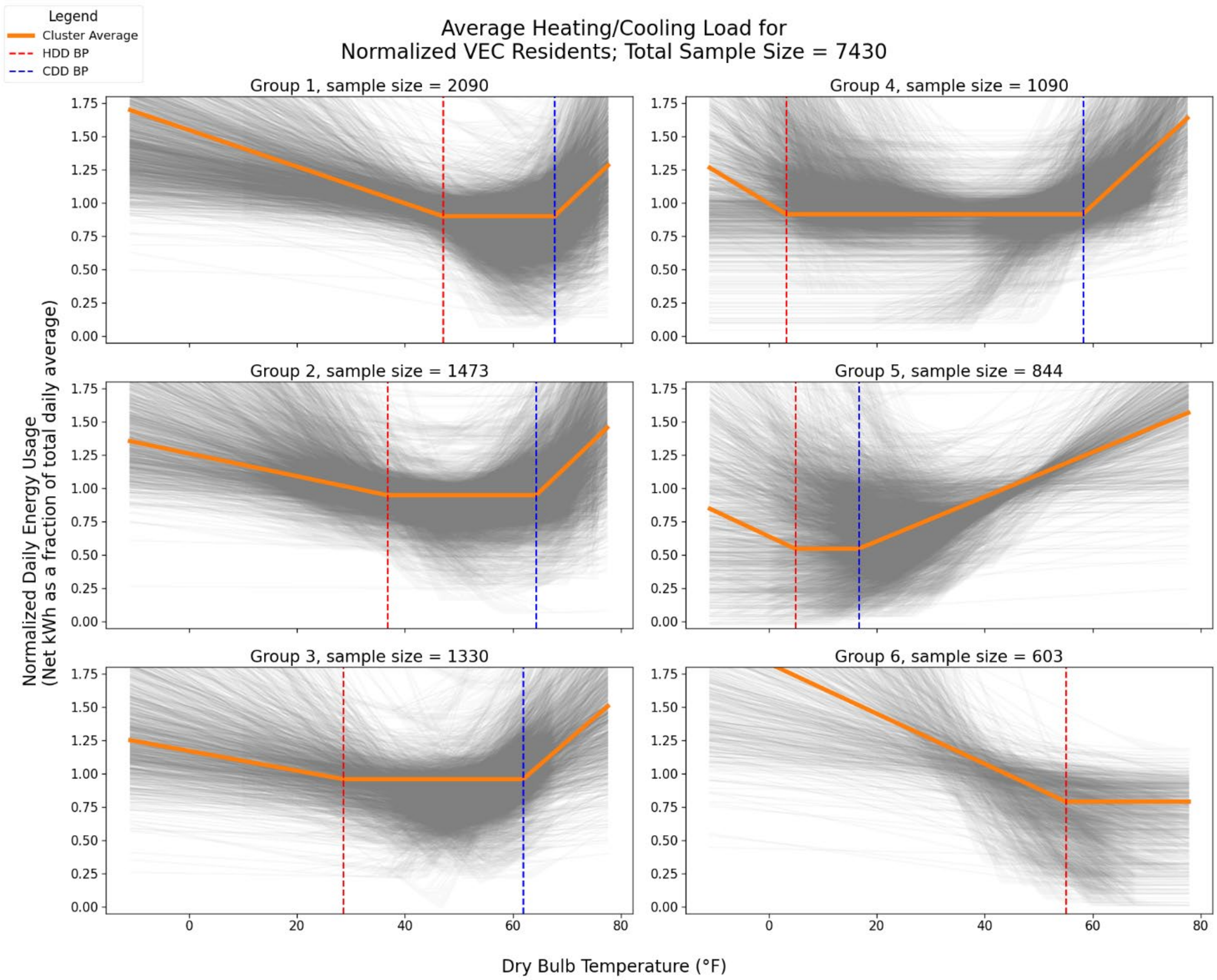

Figure 7. The average CalTRACK model shape of each cluster fitted over the plots of the individual residence model that make up each cluster. Groups 1, 2, and 3 all show varying degrees of electrical home heating efficiency with heating rising at similar rates starting at reasonable cool temperatures. Group 4 and 5 show heating turn on at extreme temperatures, suggesting an alternative heating source to electrical energy. Group 6 shows the most inefficient heating rates paired with heating balance point out of all samples.

Cluster	Percent of Total Sample	Baseline Intercept	Heating Balance Point	Cooling Balance Point	Heating Slope	Cooling Slope
1	%28.12	0.900	47.010	67.710	0.014	0.039
2	%19.83	0.948	36.791	64.291	0.009	0.039
3	%17.90	0.960	28.565	61.906	0.007	0.035
4	%14.67	0.915	3.275	58.174	0.025	0.037
5	%11.36	0.546	4.914	16.710	0.019	0.017
6	%8.12	0.791	55.027	n/a	0.019	n/a

Table 1: Parameter results for average models seen in figure 7. Also shown is the percentage of the original sample each group is composed of.

Groups 1, 2, and 3 all show a resemblance to one another. These three groups have shapes that intuitively seem characteristic of a residence that uses electricity for heating and cooling: a range of temperatures in which heating and cooling energy is not used, followed by increased energy usage once temperatures begin to get colder or warmer. Each cluster also has similar baseline intercepts, which result from normalizing the daily energy usage as a fraction of annual total. Focusing on the impact of heating, of these three typical-looking groups, Group 1 has not only the highest heating balance point, but also the largest heating slope, indicating that of these three groups, Group 1 starts heating in warmer temperatures and heats more inefficiently. While the heating balance point for Groups 2 and 3 are more similar, Group 3 has the lowest heating slope and baseline intercept out of these three. This does not necessarily mean that residential buildings in this group are more efficient at heating. Buildings in this group may have better weatherization, or even use supplemental heating from other sources that are not reflected in AMI data. Additionally, at this granularity of classification we cannot distinguish the electric consumption of heat pump components such as the usage of pump and fan systems. While we understand that Group 3 holds the potential for the most efficient buildings, the reality is that there is not enough information to confirm this.

Group 6, by contrast, has a much more extreme shape. On average, Group 6 uses no cooling energy, while having a fairly standard electrical heating load. This group's heating slope is just barely above average for our sample and its heating balance point is the highest (though by no means unreasonable for an electrical heating load). The lack of a cooling balance point and cooling slop for Group 6 indicates that no substantial increase in energy usage occurs when outdoor air temperatures rise.

In a similar fashion, Group 5 almost exclusively uses cooling energy at most temperatures, and only uses heating on the coldest days of the year. Interestingly, Group 5 has the lowest cooling slope by far, excluding Group 6; homes in this group on average have efficient cooling loads compared to others in the sample. Due to the lack of heating used on average for Group 5, this group likely uses supplemental heating as their main source, while using electrical heating at extremely low temperatures.

Finally, Group 4 has the most interesting shape in the context of this research. The most important feature to note is the large difference between the heating and cooling balance points, and at which temperatures the change in energy usage occurs. While the cooling balance point is in an appropriate range, the heating balance point is nowhere close to what a typical homeowner would select in their electrical thermostat. Because of the low heating balance point, the heating signature implies residence in this group are not using electricity as their main source of heating, but rather an alternative, non-electrical option.

Discussion

The six groups derived by our K-means clustering algorithm formed according to overarching similar features in daily energy models. Analyzing these features can provide identification criteria for groups of residences that do not use electrical energy for primary heating needs.

Groups 4 and 5 exhibit the most pronounced indications that residence in these groups do not generally use electrical energy for heating. The low value of the heating balance points in these groups reveals that their corresponding residences only use additional electrical energy at the most extreme of cold temperatures (under 3–5°F). As temperatures decrease below these heating balance points, there is a sharp rise in electrical energy use. Unless these residences are going without heat until temperatures reach 3–5°F, they must therefore be using a non-electric energy source for heating. Increased electrical energy use below the low heating balance points suggests a secondary heat source, such as a portable space heater, at extreme temperatures. This combination of low heating balance points and additional electricity usage below these balance points make Groups 4 and 5 the most likely to contain residences that use non-electric energy as their primary source of heating. However, Group 5 has some suspicious features that are unique amongst our groups; the lowest baseload usage, and an extremely low cooling balance point. Its unusual shape is difficult to interpret – are these residences really using cooling energy when it's 20°F outside? More information, like incorporating a population of known installed heat pumps, would be needed to determine if Group 5 contains residence of interest. Because of these suspicions, the team felt more confident in selecting Group 4 as the most likely group to hold participants of the unregulated fuel market.

Each of the six groups holds some special characteristics not prominent in other groups. Analyzing these characteristics can provide helpful insights beyond the suggestion of identification criteria for non-electric heating users. For example, Group 1's, 2's, and 3's similar shapes also show gradual decreases in heating balance point and heating slope between Group 1, 2, and 3. Group 3's relatively lower heating balance point and heating slope indicate that its electrical heating energy is the most efficient. For typical residences that use electricity for heating, characterizing efficiency in this way could indicate the best opportunity for efficiency improvements. For example, among these three groups, residences in Group 1 with the lowest heating efficiency would gain the most benefit from efficiency improvements, leading them to be the first priority for this work.

Groups 5 and 6 are the most atypical energy users in the project sample. Residences in these groups exclusively dedicate energy toward heating or cooling at most temperatures. One explanation for this could be attributed to wildly inefficient, but their lower baseline usage suggests that these are not typical residences like those in Groups 1–3. Instead, a more likely explanation is that Group 6 contains residences that don't use any sort of cooling system and Group 5 contains residences that are using a non-electric heating source.

The motivation of this research was to support the electrification of non-electrical heat sources, so most of this analysis was done with heating in mind, rather than cooling. Because cooling systems don't require unregulated fuels—cooling typically uses electricity as an energy source—heating and cooling systems are fundamentally different. Analyzing cooling loads is, however, another potential application of this work. For example, homes with high rates of cooling energy, such as Group 2, might be good candidates for demand management or air conditioning efficiency programs. Additionally, by performing a deeper analysis on customers that only use heat pumps, similar analysis would help distinguish heat pump performance and optimization for a given population.

Limitations

It is important to note the lack of verification for the accuracy of model classification. Because this process of characterizing residential fuel usage relies on extrapolation, the results of this analysis do not represent a complete picture of energy usage for all homes in a single group. This work gives insight into residences' electrical usage for heating and cooling, using the assumption that residences using little to no electrical heating energy are heating through non-electric means. This work does not, however, provide identification of non-electric energy sources.

Due to limited time and data availability, the research team was unable to align these results with known installations of certain heating end uses from Efficiency Vermont program data. To make more accurate classifications of various heating end uses, we suggest verification of model classification against known install data.

Conclusion

Heating and cooling energy disaggregation presents a promising opportunity to gain deeper knowledge of residential buildings that use non-electric heating. The results of this analysis show a considerable proportion of our sample using little electrical energy when lowering air temperatures occurs, implying that the main source of heating energy is unregulated fuels or wood heat rather than a more sustainable and renewable electrical source. Curated savings programs, supported by these classification methods, could address these ideal customers, whom standard outreach methods would not otherwise prioritize.

Moreover, the ability to create detailed classifications of heating signatures to capture emerging behaviors in electrical usage allows for the potential of supporting tailored savings programs that may want to target these groups. This ability to classify buildings into one of any number of characterized groups would provide energy efficiency utilities (EEUs) and program managers a higher level of insight into their customer base, as well as allow customers a more granular look at their energy usage when searching for programs for which they might qualify.